Automated Neuropsychological Assessment Metrics: Repeated Assessment with Two Military Samples

Stephanie E. Eonta, Walter Carr, Joseph J. McArdle, Jason M. Kain, Charmaine Tate, Nancy J. Wesensten, Jacob N. Norris, Thomas J. Balkin, and Gary H. Kamimori

Eonta SE, Carr W, McArdle JJ, Kain JM, Tate C, Wesensten NJ, Norris JN, Balkin TJ, Kamimori GH. ***Automated Neuropsychological Assessment Metrics: repeated assessment with two military samples.*** **Aviat Space Environ Med 2011; 82:34–9.**

Introduction: U.S. military troops deploying to war zones are currently administered the Automated Neuropsychological Assessment Metrics (ANAM4) Traumatic Brain Injury (TBI) Battery to establish individual neurocognitive performance baselines. In part, the utility of the ANAM4 TBI Battery baseline measurement depends on test-retest reliability of this instrument. The purpose of this report was to evaluate performance following multiple administrations of the ANAM4 TBI Battery: does performance in a repeated measures paradigm constitute a stable, interpretable indication of baseline neurocognitive ability? ***Methods:*** The data presented here are from the ANAM4 TBI Battery administered four times to a group of U.S. Marines in Study 1 and eight times to a group of New Zealand Defence Force personnel in Study 2. ***Results:*** The results show practice effect in five of six performance subtests in both Study 1 and Study 2. ***Discussion:*** Results are consistent with expectations that multiple test sessions are required to reach stable performance on some computerized tasks. These results have implications for taking ANAM4 TBI Battery practice effects into account in test administration and in data interpretation.

Keywords: neuropsychological assessment, computer-based testing, practice effects, repeated measures.

THE AUTOMATED Neuropsychological Assessment Metrics Traumatic Brain Injury Battery [ANAM4 TBI Battery (1,18,24)] is currently being administered to U.S. military troops, per law passed by Congress, prior to deployment, as a baseline measure of neurocognitive performance (2). Subtests in the ANAM4 TBI Battery were selected for their presumed sensitivity to cognitive deficits commonly associated with head trauma [e.g., increased reaction time (17,24)]. The sensitivity of the various ANAM4 subtests for detecting neurocognitive deficits has been tested in a variety of military and nonmilitary contexts, including those involving sports-related concussion and exposure to radiation, high altitude, undersea conditions, and toxins (10,14,20). The putative advantages of ANAM4 (compared to other neuropsychological assessment batteries) are that it is a standardized, computerized, self-guided test platform with multiple alternate forms—a design that makes it specifically suitable for repeated administration. Despite the known prevalence and well-understood implications of practice effects, research on the effects of multiple administrations of the same neuropsychological test, especially the ANAM4 TBI Battery, is sparse (16,23).

Repeated assessment is commonly used to track the progression of disease or injuries (7) and baseline measurements can improve the sensitivity of such assessments. Performance improvement on repeated neurocognitive tests often occurs because individuals refine their strategies over repeated exposures to a test and, consequently, improve their performance (5,21). Practice effects vary as a function of number of administrations, time between administrations, and test complexity (6,15,21). If practice effects are not taken into account, then results of subsequent assessments may be confounded and, thus, misinterpreted, e.g., as disease remission or intervention-related improvement (6,15,22).

With respect to the ANAM TBI Battery, there is one technical report published in which practice effects were assessed (8). In that report it was concluded that practice effects did not persist beyond the third administration. However, neuropsychological assessments similar to the current ANAM4 TBI Battery have shown significant practice effects after three administrations (9,28). ANAM batteries have a history of variance in form across users and can evolve over time. Thus, the Bendetto et al. (8) results may not generalize to the current ANAM4 TBI Battery version. The current ANAM4 TBI Battery is comprised of some subtests presumed to be more difficult than the one used by Bendetto et al. because they require visuospatial information recall without rehearsal (3,4).

The purpose of this study was to characterize and evaluate our use of repeated testing with the ANAM4 TBI Battery to mitigate practice effects and achieve a relatively stable level of performance for samples of military personnel. Our criterion for practice effect was significant performance improvement between the first

From the Naval Medical Research Center, Silver Spring, MD; Walter Reed Army Institute of Research, Silver Spring, MD; the Naval Health Research Center, Silver Spring, MD; and the New Zealand Defence Force, Auckland, New Zealand.

This manuscript was received for review in March 2010. It was accepted for publication in October 2010.

Address correspondence and reprint requests to: Stephanie Eonta, Operational & Undersea Medicine/Neurotrauma, Naval Medical Research Center, 503 Robert Grant Ave., Silver Spring, MD 20910-7500; stephanie.eonta@med.navy.mil.

DOI: 10.3357/ASEM.2799.2011

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JAN 2011		**00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Automated Neuropsychological Assessment Metrics: Repeated Assessment with Two Military samples	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Medical Research Center, , ,Silver Spring,MD**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

This work was supported and funded by work unit number 602702E.0000.000A0811. The study protocols were approved by the Naval Medical Research Center Institutional Review Board in compliance with all applicable Federal regulations governing the protection of human subjects. From the Naval Medical Research Center, Silver Spring, MD; Walter Reed Army Institute of Research, Silver Spring, MD; the Naval Health Research Center, Silver Spring, MD; and the New Zealand Defence Force, Auckland, New Zealand.

14. ABSTRACT

Introduction: U.S. military troops deploying to war zones are currently administered the Automated Neuropsychological Assessment Metrics (ANAM4) Traumatic Brain Injury (TBI) Battery to establish individual neurocognitive performance baselines. In part, the utility of the ANAM4 TBI Battery baseline measurement depends on test-retest reliability of this instrument. The purpose of this report was to evaluate performance following multiple administrations of the ANAM4 TBI Battery: does performance in a repeated measures paradigm constitute a stable, interpretable indication of baseline neurocognitive ability? Methods: The data presented here are from the ANAM4 TBI Battery administered four times to a group of U.S. Marines in Study 1 and eight times to a group of New Zealand Defence Force personnel in Study 2. Results: The results show practice effect in fi ve of six performance subtests in both Study 1 and Study 2. Discussion: Results are consistent with expectations that multiple test sessions are required to reach stable performance on some computerized tasks. These results have implications for taking ANAM4 TBI Battery practice effects into account in test administration and in data interpretation.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Public Release**	**6**	

and the last available timepoints. Our criterion for 'stable performance' was non-significant change in performance between the last two available time points. Given that these subjects are operational military personnel tested outside of the laboratory environment, a key criterion in the design of the research protocols and the repeated testing was feasibility—there were a limited number of available time points for test administration. We included as many time points as were feasible. Data were drawn from two studies: in Study 1, the ANAM4 TBI Battery was administered four times; in Study 2, the ANAM4 TBI Battery was administered more frequently (eight times). We expected practice effects in our use of multiple sessions, given other reports of practice effects in repeated assessments with neurocognitive testing (6,23), especially computer-based neurocognitive testing (26), including forms of ANAM subtests (not the TBI battery) (25).

METHODS

Study 1

Subjects

Subjects were 38 active duty male U.S. Marines (mean age = 25, range = 19 to 35) from a training facility in Marine Corps Base Quantico. Educational level was relatively homogenous across subjects. All subjects had completed high school education, with only 2 of the 38 subjects completing a higher degree.

Procedure

Subjects consented to participate in this study during an informed consent meeting. Per IRB guidance, the informed consent meetings were conducted individually. Subjects completed the ANAM4 TBI Battery in groups on individually assigned laptop PCs. Testing was in a classroom-like setting with 3′ W × 2′ H cardboard dividers surrounding each workspace. Each subject was administered the battery four times (Sessions 1-4) across 4 d. Sessions 1 and 2 were back-to-back on the first day (Saturday or Sunday) and Sessions 3 and 4 were on the following Monday and Tuesday, respectively. Subjects completed each administration of the battery at approximately 1600.

Instrument

The ANAM4 TBI Battery is a PC-based neuropsychological assessment consisting of two subjective subtests, Sleep Scale (SLP) and Mood Scale (MOO) (in order), followed by six performance subtests: Simple Reaction Time (SRT), Code Substitution (CDS), Procedural Reaction Time (PRO), Mathematical Processing (MTH), Matching To Sample (M2S), and Code Substitution Delayed (CDD) (in order) (11). For performance subtests, stimuli are presented visually and responses are two-choice forced choice responses using left and right buttons on a computer mouse. These eight subtests are described in additional detail below.

Sleepiness scale: The SLP is a self-assessment of the subject's sleepiness/fatigue state. Seven different statements of alertness/sleepiness are presented to the subject ranging from, "Feeling very alert, wide awake, and energetic" to "Very sleepy and cannot stay awake much longer."

Mood scale: The MOO is a self-assessment of the user's mood state in seven categories: Vigor (high energy level), Happiness (positive disposition), Depression (dysphoria), Anger (negative disposition), Fatigue (low energy level), Anxiety (anxiety level), and Restlessness (motor agitation). A series of adjectives, each contributing to one of the mood categories, is presented to the subject. Subjects are instructed to indicate on a scale of 0 to 6 the number that best represents their current state, with "0" as "Not at all" to "6" as "Very Much."

Simple reaction time: The subject clicks the left mouse button (single-button response) when an asterisk stimulus is presented on the screen. This stimulus is presented at different intervals for 40 trials and reaction time for each trial is recorded. This subtest assesses reaction time. Note that the first test (SRT) is administered twice in the ANAM4 TBI Battery, as the first and as the last performance subtest. In the variant of the ANAM4 TBI Battery used here the second administration of the SRT (SR2) included an additional 60 trials—otherwise, the subtests in the battery reported here are identical to those in the ANAM4 TBI Battery. Given the additional trials in SR2, the results for SR2 are not reported here.

Code substitution: A static display of digits 1 through 9 appears in a row at the top of the screen with a unique symbol (e.g., δ, √, ↕, ≡, æ, ╣, Ω, ¥, ◄) above each digit. A series of 72 probes appears at the bottom of the screen, each showing a pairing of a single digit and symbol in the same fashion as the static display at the top of the screen. The subject uses the left mouse button to indicate if the pairing in the probe matches a pairing in the static display above and the right mouse button if the pairing in the probe does not match a pairing in the static display above. Subjects are informed that the static display will be referenced in a later task, but the display will not be represented. This subtest assesses visual search, sustained attention, and encoding.

Procedural reaction time: A series of single digits (2, 3, 4, or 5) is presented in 32 trials. The subject uses the left mouse button to indicate the digit is "low" (2 or 3) or the right mouse button to indicate the digit is "high" (4 or 5). This subtest assesses reaction time and processing efficiency associated with following a simple set of mapping rules.

Mathematical processing: A series of 3 single-digit operator arithmetic mathematical equations (e.g., "3 + 4 - 1") is presented in 20 trials. The subject uses the left mouse button to indicate the answer is less than 5 or the right mouse button if the answer is greater than 5. This subtest assesses basic computational skills, concentration, and working memory.

Matching to sample: A series of 4 × 4 matrices with cells in a 2-colored pattern appears in 20 trials. Following each stimulus, a pair of slightly different 4 × 4 matrices appears side-by-side. The subject uses the left or right mouse button to indicate which matrix in the pair

matches the previous stimulus. This subtest assesses spatial processing and visuo-spatial working memory.

Code substitution delayed: A series of 36 probes appears in the same fashion as the CDS subtest. The subject responds in the same fashion as in the CDS subtest using memory of the static display from the CDS subtest, presented approximately 10 min before and not represented.

In our initial administration of the ANAM4 TBI Battery there were practice trials administered as part of the battery at the beginning of each of the performance subtests with the objective of familiarizing the subject with the nature of the subtests. These practice trials were not used in subsequent administrations. ANAM was specifically designed to meet the needs of repeated-measures testing in assessment of neuropsychological functioning. "The system has a pseudo-randomization procedure that permits creation of multiple forms from item sets, thus allowing tests to be used for performance monitoring and in repeated-measures designs" (11). Specifically, in multiple administrations of the ANAM4 TBI Battery the stimuli for the performance subtests are varied according to session number when possible. The asterisk stimulus in SRT and the digits 1-9 in CDS and CDD are not replaced or reordered and so are presented in the same fashion for all administrations of those subtests. These alternate forms of ANAM, generated by the pseudo-randomized selection of comparable test items that are then presented in the same test paradigm, are widely reported as equivalent in published research (25,30).

Dependent Variables

Principal dependent variables for all performance subtests are reaction time and accuracy. In typical use of ANAM4, these two dependent variables are combined into a single outcome measure of "throughput" (27). Throughput is derived from percent correct divided by mean reaction time and, conceptually, is a speed-accuracy product reflecting performance across both dependent variables. For clarity, throughput is the variable presented for the performance subtests in this report. The subjective response subtests, SLP and MOO, do not have an accuracy component so those results are reported here as reaction time rather than throughput.

Data Exclusion Criteria

To rule out potential systematic confounds in our analysis of repeated administration of the ANAM4 TBI Battery in this sample, we applied medical and scheduling criteria to mitigate likely competing sources of variance. Although 38 Marines completed this study, 4 were excluded from this analysis based on medical conditions, specifically pre-existing medical conditions that would affect the brain (e.g., history of diagnosed TBI), yielding a sample size of 34.

Statistical Analyses

Session means for throughput were compared using one-way repeated measures ANOVA. Separate analyses were performed for each of the ANAM subtests. Session was treated as a repeated variable. Huynh-Feldt epsilon adjustments were applied to all F statistics to correct for potential covariation between the experimental conditions introduced by the repeated measures design. A planned comparison (*t*-test) for performance improvement was conducted between the first session and the last session (Sessions 1 v. 4) for each subtest. A planned comparison for stable performance was conducted between the last two sessions (Sessions 3 v. 4). The planned *t*-test comparison for performance improvement was performed only when the F-test associated with a specific hypothesis yielded an effect significant at $P \leq 0.05$. Bonferroni corrections were applied to the planned comparisons to maintain a family-wise α of 0.05. Analyses were performed with SAS and Matlab R2009b.

Study 2

Subjects

Subjects were 21 active-duty male New Zealand Defence Force soldiers (mean age = 28, range = 22 to 43) from a New Zealand training facility. Educational level was relatively homogenous across subjects. All subjects had completed the New Zealand equivalent of high school education and 3 of the 21 subjects had completed a higher degree. Two subjects did not report education level.

Procedure

Subjects attended an informed consent briefing and then, per IRB guidance, were given a 24-h period before granting consent to participate in the study. Subjects completed the ANAM4 TBI Battery in groups on assigned laptop PCs. Testing was in a classroom-like setting with 3' W × 2' H cardboard dividers surrounding the workspace. Each subject was administered the ANAM4 TBI Battery eight times (Sessions 1-8) over the course of 5 d. Sessions 1-3 were back-to-back-to-back in the late morning on the first day, Sessions 4-5 and Sessions 6-7 were morning and afternoon, respectively, on the next 2 d. Session 8 was morning of the following day. The Instrument (ANAM4 TBI Battery) and dependent variables (throughput and reaction time) were identical to Study 1.

Data Exclusion Criteria

For analyses, two subjects were excluded due to incomplete data and seven subjects were excluded according to Study 1 data exclusion criteria. Specifically, seven were excluded due to a difference in testing schedule from the primary group. This yielded a sample size of 12.

Statistical Analyses

Analysis approach was identical to Study 1, except that planned comparisons for performance improvement were between Sessions 1 and 8 and planned comparisons for stable performance were between Sessions 7 and 8.

RESULTS

Performance improvement was seen across test administration sessions in both samples, but this improvement was not seen for all subtests in both samples. Performance across all time points is presented in **Fig. 1** and *t*-test comparisons are presented in **Tables I and II**. Improvements across repeated administrations were most clearly seen in both samples for the SRT subtest. MTH also showed improvement in both samples, but the pattern of improvement was not as strong as in SRT. Improvement was seen in PRO for Study 1 and in M2S, CDS, and CDD for Study 2. No other performance improvement reached statistical significance in our criterion. The criterion for stable performance at the end of repeated administration (i.e., non-significant change in performance between the last two available time points) was met for SRT, PRO, M2S, MTH, and CDS in both samples, but for CDD only in Study 1. Interestingly, CDD showed continued improvement in performance between Session 7 and Session 8 in Study 2.

In Fig. 1 there is an apparent decline in performance between Sessions 1 and 2 for the CDS subtest. That decline is statistically significant in both Study 1 and Study 2 [$t(33) = 3.852$, $P = 0.001$ and $t(11) = 3.962$ $P = 0.002$, respectively]. This comparison was not part of the planned comparisons, but was remarkable given that it seemed parallel in both samples.

In general, performance improvement was not associated with variance from age or education. Among the 24 correlations available for the 6 subtests, 2 planned comparisons, and the 2 studies, only 3 reached statistical significance (Fisher's r to Z). In Study 1, age was correlated with performance stability in SRT and MTH ($r = -0.376$, $P = 0.028$ and $r = 0.373$, $P = 0.029$, respectively). In Study 2, education was correlated with performance stability in PRO ($r = -0.664$, $P = 0.016$). Given that these correlations were infrequent and were inconsistent in direction, they were not considered meaningful for this report.

The subjective subtest MOO showed faster reaction times between the first session and last session for all subscales and in both samples (*P*-values for all *t*-statistics less than 0.01). SLP showed the same pattern for Study 1, but in Study 2 there was no change in reaction time between the first and last sessions. These decreases in reaction time for MOO and SLP occurred almost entirely between Sessions 1 and 2. The one exception was that the MOO Anger subscale showed a decrease in reaction time between Session 3 and Session 4 in Study 1 [$t(33) = 2.489$, $P = 0.026$]. Otherwise, stable performance between the last two available time points was observed.

DISCUSSION

The purpose of this study was to characterize and evaluate our use of repeated testing with the ANAM4 TBI Battery to mitigate practice effects and achieve a relatively stable level of performance. Our targeted criterion allowed for an overall improvement across all time points in ANAM performance subtests, but requires no change (or non-significant improvement) between last two sessions. Our use of the ANAM4 TBI Battery in these two military samples met this objective, with one potential exception, CDD for Study 2. We regard this as a "potential" exception because the final performance (Session 8) on this subtest in that sample may have been at asymptote. The evidence for performance level on this subtest being at asymptote is from mean comparison to the asymptotic performance in the corresponding Study 1 subtest (see Fig. 1); however, asymptotic performance in Study 2 was not supported using our stated criterion.

The subjective subtests, SLP and MOO, are reflections of state rather than performance and, thus, would not be

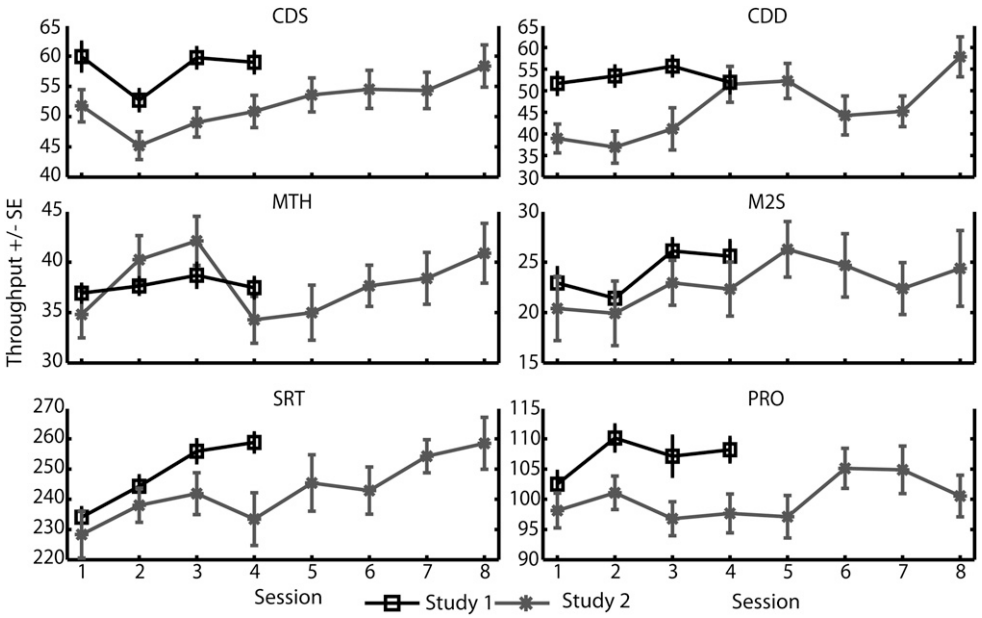

Fig. 1. Means for throughput (± SE) on ANAM4 subtests SRT, CDS, PRO, M2S, MTH, and CDD from both Study 1 and Study 2. Throughput units are correct responses per minute of responding (27). For Study 1 the final session was Session 4. For Study 2 the final session was Session 8.

TABLE I. STUDY 1 TIME POINT COMPARISONS OF THROUGHPUT FOR EACH PERFORMANCE SUBTEST.

Subtest	F-test	*t*-test (S1 vs. S4)	*t*-test (S4 vs. S3)
SRT	$F(3,99) = 20.32$, $P = 0.0001$, $\epsilon = 0.9137$	$t(33) = -6.94803$, $P = 0.0002$	$t(33) = 0.81939$, $P = 0.829$
PRO	$F(3,99) = 3.77$, $P = 0.0195$, $\epsilon = 0.8254$	$t(33) = -2.4136$, $P = 0.0352$	$t(33) = 0.457612$, $P = 1$
M2S	$F(3,99) = 0.5$, $P = 0.6834$, $\epsilon = 1.0463$		$t(33) = -0.82963$, $P = 0.8174$
MTH	$F(3,99) = 17.98$, $P < 0.0001$, $\epsilon = 0.9606$	$t(33) = -3.58988$, $P = 0.001$	$t(33) = -0.70266$, $P = 0.9678$
CDS	$F(3,99) = 9.88$, $P = 0.0005$, $\epsilon = 0.5484$	$t(33) = 0.603715$, $P = 1$	$t(33) = -0.4847$, $P = 1$
CDD	$F(3,99) = 1.28$, $P = 0.2864$, $\epsilon = 1.0095$		$t(33) = -1.62715$, $P = 0.2138$

S1 = Session 1; S4 = Session 4; S3 = Session 3; SRT = Simple Reaction Time; PRO = Procedural Reaction Time; M2S = Matching To Sample; MTH = Mathematical Processing; CDS = Code Substitution; and CDD = Code Substitution Delayed.

expected to exhibit practice effects with repeated measures. We examined reaction time here because familiarity and accommodation to the computer-based format should be expected to reflect in increased speed of responding. That increased speed of responding would consequently account for a degree of improvement on the performance subtests as well. Our expectation was supported in the results here. We did not expect this practice effect to be observed in the final two sessions of these subjective subscales, but that did occur for the MOO Anger subscale in Study 1. However, that increase in speed of response was the only such result in the set of 16 planned comparisons.

The practice effects reported here have implications for researchers and practitioners interpreting pre- and post-test results in the evaluation of neuropsychological change. The confirmation of practice effects in ANAM4 is an important substantiation for the study design used in an ongoing program of research on bio-effects from repeated exposure to low-level blast (12). This report was not an effort to yield careful scaling of the magnitude of practice effects. Guidance on the extent practice effects relate to interpretability of test battery results would be more appropriate for a study specifically designed for that purpose.

In other observations from these results, there are suggestions of systematic influence on performance beyond that of the repeated administration. Specifically, there are suggestions in the data of day-to-day variance and within day variance as well as an interference effect in back-to-back administration of CDS. Day-to-day and within day effects might be exhibited in SRT for Study 2. In Study 2, Sessions 1-3 were administered back-to-back-to-back and Session 4 was administered on a following day. Massing Sessions 1-3 followed by an overnight intersession interval before Session 4 may account for the mean increase in SRT throughput for Sessions 1-3 and drop-off for Session 4. The effect of session massing also may be evidenced in the change between SRT Sessions 5 and 6 and, for the MTH and CDD subtests, in the transition between Sessions 3 and 4 and between Sessions 5 and 6. This day-to-day variance may not always result in performance decrease (e.g., throughput increase between Sessions 3 and 4 in CDD for Study 2). Within day effects (i.e., morning vs. afternoon) might be exhibited in the throughput pattern seen between Sessions 4 and 5 and between Sessions 6 and 7 of SRT for Study 2, although this pattern was not reflected elsewhere in these data.

The effect of proactive interference following repeated administration using similar stimuli may account for the pattern of results seen in the CDS subtest (Fig. 1) (13,29). Both Study 1 and Study 2 show a marked decline in throughput between Session 1 and Session 2, which were administered back-to-back. This phenomenon, a performance decrease despite immediately preceding practice, may be explained by proactive interference resulting from the pairing of digits 1-9 to a set of symbols in Session 1 and a different, but similar set of symbols in Session 2. Interestingly, this interference seems to be present predominantly between Sessions 1 and 2, and not between other sessions when administered within the same day (e.g., Sessions 4 and 5 in Study 2) or even when administered back-to-back (e.g., Sessions 2 and 3 in Study 2). Furthermore, this interference effect does not seem to occur in CDD, which uses the same stimuli as CDS. This may be the result of differentiation aspects of perceptual learning (for a review see 19). Jointly, proactive interference and perceptual learning should be taken into account when this subtest

TABLE II. STUDY 2 TIME POINT COMPARISONS OF THROUGHPUT FOR EACH PERFORMANCE SUBTEST.

Subtest	F-test	*t*-test (S1 vs. S8)	*t*-test (S7 vs. S8)
SRT	$F(7,77) = 4.84$, $P = 0.0011$, $\epsilon = 0.6964$	$t(11) = -4.67606$, $P = 0.0002$	$t(11) = -0.66013$, $P = 1$
PRO	$F(7,77) = 1.9$, $P = 0.0995$, $\epsilon = 0.8073$		$t(11) = 1.257313$, $P = 0.4248$
M2S	$F(7,77) = 2.69$, $P = 0.0238$, $\epsilon = 0.8158$	$t(11) = -2.33579$, $P = 0.0442$	$t(11) = -0.95726$, $P = 0.6828$
MTH	$F(7,77) = 6.1$, $P < 0.0001$, $\epsilon = 1.0487$	$t(11) = -3.22975$, $P = 0.0036$	$t(11) = -1.6213$, $P = 0.218$
CDS	$F(7,77) = 5.87$, $P = 0.0018$, $\epsilon = 0.4647$	$t(11) = -2.8269$, $P = 0.012$	$t(11) = -1.73745$, $P = 0.1726$
CDD	$F(7,77) = 8.24$, $P < 0.0001$, $\epsilon = 0.8356$	$t(11) = -5.2882$, $P = 0.0002$	$t(11) = -3.52702$, $P = 0.0014$

S1 = Session 1; S8 = Session 8; S7 = Session 7; SRT = Simple Reaction Time; PRO = Procedural Reaction Time; M2S = Matching To Sample; MTH = Mathematical Processing; CDS = Code Substitution; and CDD = Code Substitution Delayed.

is administered in a repeated fashion, including when this subtest is administered a second time following a subject's initial performance that falls below the ANAM4 standard criterion of 56% accuracy (11). When a subject is taking CDS a second time following below criterion performance, it may be that limited encoding of the stimuli pairings in the initial test occurred and, consequently, there is no interference in the second test. However, that is an assumption and does not preclude some level of caution when administering CDS a second time with the same stimuli in an alternate form. This potential for interference is mitigated if the second administration is reset to use the same stimuli in the same form as used in the initial session.

In conclusion, a stable level of performance was achieved in these repeated administrations of the ANAM4 TBI Battery, yielding a performance baseline that would be useful in subsequent evaluation for change effects. The performance improvement observed prior to that stable level of performance suggests that use of the ANAM4 TBI Battery with a single repeated administration may yield an underestimate or an inaccurate estimate of change effects. This suggested caution is notable in that the practice effect reported here in component subtests of the ANAM4 TBI Battery was based on two separate studies with varying test administration paradigms. A quantitative model of the performance improvement effects, however, is outside the scope of this report and warrants studies designed for that purpose.

ACKNOWLEDGMENTS

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of the Army, Department of Defense, New Zealand Defence Force, nor the U.S. Government.

This work was supported and funded by work unit number 602702E.0000.000A0811. The study protocols were approved by the Naval Medical Research Center Institutional Review Board in compliance with all applicable Federal regulations governing the protection of human subjects.

Authors and affiliations: Stephanie E. Eonta, B.S., B.A., Walter Carr, Ph.D., and Joseph J. McArdle, Ph.D., Naval Medical Research Center, Silver Spring, MD; Jason M. Kain, B.S., Nancy J. Wesensten, Ph.D., and Gary Kamimori, Ph.D., Walter Reed Army Institute of Research, Silver Springs, MD; Charmaine Tate, M.D., New Zealand Defence Force, Auckland, New Zealand; and Jacob N. Norris, Ph.D., Warfighter Performance, Naval Health Research Center, San Diego, CA.

REFERENCES

1. Automated Neuropsychological Assessment Metrics, 4 ed. Norman, OK: C-SHOP; 2007.
2. National Defense Authorization Act for Fiscal Year 2008, Public Law 110-181. HR 1585, Sect. 1618, 110th Congress (2008).
3. Baddeley A, Chincotta D, Adlam A. Working memory and the control of action: evidence from task switching. J Exp Psychol Gen 2001; 130:641–57.
4. Baddeley AD. Is working memory still working? Am Psychol 2001; 56:851–64.
5. Basso MR, Bornstein RA, Lang JM. Practice effects on commonly used measures of executive function across twelve months. Clin Neuropsychol 1999; 13:283–92.
6. Basso MR, Carona FD, Lowery N, Axelrod BN. Practice effects on the WAIS-III across 3- and 6-month intervals. Clin Neuropsychol 2002; 16:57–63.
7. Beglinger LJ, Gaydos B, Tangphao-Daniels O, Duff K, Kareken DA, et al. Practice effects and the use of alternate forms in serial neuropsychological testing. Arch Clin Neuropsychol 2005; 20:517–29.
8. Bendetto J, Harris W, Goernert P. Automated Neuropsychological Assessment Metrics (ANAM) performance stability during repeated cognitive assessments. Mankato, MN: Mankato State University; 1995. Technical Report EPRL-TR-95-02.
9. Benedict RH, Zgaljardic DJ. Practice effects during repeated administrations of memory tests with and without alternate forms. J Clin Exp Neuropsychol 1998; 20:339–52.
10. Bleiberg J, Cernich AN, Cameron K, Sun WY, Peck K, et al. Duration of cognitive impairment after sports concussion. Neurosurgery 2004; 54:1073–8.
11. C-Shop. ANAM4 TBI: user manual. Norman, OK: Center for the Study of Human Operator Performance, University of Oklahoma; 2007.
12. Carr W, Stone J, Young L, Walilko T, Parish R, et al. Breacher injury study: neurocognitive testing results. [Poster Abstract.] J Neurotrauma 2010; 27(5, Suppl.):A29.
13. Ceraso J. The interference theory of forgetting. Sci Am 1967; 217:117–24.
14. Cernich A, Reeves D, Sun WY, Bleiberg J. Automated Neuropsychological Assessment Metrics sports medicine battery. Arch Clin Neuropsychol 2007; 22(Suppl. 1):S101–14.
15. Collie A, Maruff P, Darby DG, Mcstephen M. The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. J Int Neuropsychol Soc 2003; 9:419–28.
16. Duff K, Westervelt HJ, Mccaffrey RJ, Haase RF. Practice effects, test-retest stability, and dual baseline assessments with the California Verbal Learning Test in an HIV sample. Arch Clin Neuropsychol 2001; 16:461–76.
17. Falleti MG, Maruff P, Collie A, Darby DG. Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. J Clin Exp Neuropsychol 2006; 28:1095–112.
18. Friedl KE, Grate SJ, Proctor SP, Ness JW, Lukey BJ, Kane RL. Army research needs for automated neuropsychological tests: monitoring soldier health and performance status. Arch Clin Neuropsychol 2007 Feb 22(Suppl. 1):S7–14.
19. Goldstone RL. Perceptual learning. Annu Rev Psychol 1998; 49:585–612.
20. Lowe M, Harris W, Kane RL, Banderet L, Levinson D, Reeves D. Neuropsychological assessment in extreme environments. Arch Clin Neuropsychol 2007; 22(Suppl. 1):S89–99.
21. McCaffrey RJ, Ortega A, Orsillo SM, Nelles WB, Haase RF. Practice effects in repeated neuropsychological assessments. Clin Neuropsychol 1992; 6:32–42.
22. Mccaffrey RJ, Lynch JK. A methodological review of "method skeptic" reports. Neuropsychol Rev 1992; 3:235–48.
23. Mccaffrey RJ, Ortega A, Haase RF. Effects of repeated neuropsychological assessments. Arch Clin Neuropsychol 1993; 8:519–24.
24. Reeves DL, Winter KP, Bleiberg J, Kane RL. ANAM genogram: historical perspectives, description, and current endeavors. Arch Clin Neuropsychol 2007; 22(Suppl. 1):S15–37.
25. Roebuck-Spencer T, Sun W, Cernich AN, Farmer K, Bleiberg J. Assessing change with the Automated Neuropsychological Assessment Metrics (ANAM): issues and challenges. Arch Clin Neuropsychol 2007; 22(Suppl. 1):S79–87.
26. Schranz S, Osterode W. [Training effects in computer-assisted psychological performance tests]. Wien Klin Wochenschr 2009; 121:405–12.
27. Thorne DR. Throughput: a simple performance index with desirable characteristics. Behav Res Methods 2006; 38:569–73.
28. Watson FL, Pasteur MAL, Healy DT, Hughs EA. Nine parallel versions of four memory tests: an assessment of form equivalence and the effects of practice on performance. Human Psychopharmacology: Clinical & Experimental 1994; 9:51–61.
29. Wickens DD. Characteristics of word encoding. In: Melton AW, Martin E, eds. Coding processes in human memory. Washington, DC: Winston & Sons; 1972.
30. Wilken JA, Sullivan CL, Lewandowski A, Kane RL. The use of ANAM to assess the side-effect profiles and efficacy of medication. Arch Clin Neuropsychol 2007; 22(Suppl. 1):S127–33.